Modeling spatial covariation of summer temperatures and bio-indicators in an Arctic coastal area

Lennart Nilsen1, Daniel Joly2,*, Arve Elvebakk3, Thierry Brossard2

1Department of Arctic and Marine Biology and 3Tromsø University Museum, University of Tromsø, 9037 Tromsø, Norway
2Laboratoire THEMA, Université de Franche-Comté, 25030 Besançon Cedex, France

ABSTRACT: In the Arctic, temperature is a major environmental factor controlling the occurrence, abundance and distribution of plants at regional and local scales alike. This means that statistical models of temperature distribution can predict the distribution of plant species or communities. Conversely, certain plant taxa make good bio-indicators reflecting long-term thermal conditions in a given habitat. Both these assumptions were taken into account when modelling the spatial relationship between plants and temperature. This work continues a previous preliminary 1 yr study based on correlations between a plant-based index of thermophily (It) and different synthetic temperature distribution characteristics. To strengthen confidence in the results and conclusions, more temperature data were collected through a field campaign conducted over a further 5 yr period (2001 to 2005). The goals here were (1) to establish an accurate interpolation model capable of restoring, at local scale, the continuous summertime thermal raster surface, (2) to evaluate the capacity of the temperature values obtained from the model to predict the distribution of It, and (3) to extrapolate temperature surfaces from this It. The results show that the mutual predictive power between temperature and It is satisfactory and that the model can be applied to neighbouring areas, although the present study area is too small to define the geographical limits of extrapolation. This predictive power declines where local landscape structures are heterogeneous. Correlations between It and growing degree day (GDD) values derived from the modelled temperature layers were systematically analysed in order to identify conditions in which this covariation works or fails.

KEY WORDS: Temperature · Vegetation · Bio-indicators · Regression analysis · Svalbard

1. INTRODUCTION

The Intergovernmental Panel on Climate Change (IPCC) claims the Arctic may be greatly affected by global warming (IPCC 2007). The Arctic climate is currently warming rapidly and much larger changes are projected over the next century (Arctic Climate Impact Assessment 2004). IPCC climate models predict that the average temperature of land areas north of 65° will rise by more than twice the average global temperature rise between the 19th and the 21st centuries (IPCC 2007). Effects such as melting permafrost (Fronzek et al. 2009) and glaciers (Grabherr et al. 2005, Nakatsubo et al. 2010), shrinking Arctic sea ice (Miller et al. 2010), longer growing seasons (Karlsen et al. 2007), and expanding scrub cover in the tundra (Stow et al. 2004, Elmendorf et al. 2012, Epstein et al. 2012) have already been reported. Arctic tree lines have moved northwards into the tundra (Olthof & Pouliot 2010) and alpine tree lines upwards into the mountains (Danby & Hik 2007). The Zackenberg Research Station in northeast Greenland reports a warming of 2.5°C in the mean annual temperature since 1991, and an average increase in precipitation of 1.9 mm yr⁻¹ over the last 50 yr (Melttof et al. 2008), so there is now ~100 mm yr⁻¹ more precipitation than in the mid 1950s. There are also signs of climate warming in Svalbard: average air temperature has in-
creased by 0.22°C per decade from 1912 to 2007, permafrost temperatures have increased by 0.35°C over the last 15 to 20 years, and winter ice accumulation has decreased while summertime glacial melting has increased (Hansen 2010).

There are obviously close relationships between temperature and the spatial dynamics of plants, however a better understanding of the mechanisms involved is essential for assessing the impacts of global change. Little is known about how Svalbard’s vegetation has reacted to these changes, or about the magnitude of the response. There are no long-term time-series data on vegetation responses for the Svalbard archipelago. However, a field experiment has been running since 2006 in Adventdalen, central Spitsbergen (Cooper et al. 2011), and a straightforward vegetation monitoring project has been established in Endalen (Aarrestad et al. 2010), a tributary valley to Adventdalen. Many plant species are good bio-indicators, i.e. their occurrence and abundance reflect certain environmental characteristics such as rock substrate (Elvebakk 1982), soil and air moisture content (Gignac 2001), snow cover duration (Scott & Rouse 1995, Beck et al. 2005, Odland & Munkejord 2008), or air temperature (Schroder et al. 2006). Temperature in particular exerts strong effects on almost all aspects of Arctic ecosystems, including soil stability, moisture and nutrient availability, and so may effectively limit the presence of an individual species in the ecosystem (Walker 1995). Although it may be that plants adapt to climate change in the long term, the expected effect of repeated higher temperatures over a series of years is an adjustment in their ranges. Before any predictions can be made about the extension of plant ranges, we first need to know more about the present relationship between plants and climate. This is the challenge addressed in this study.

This relationship is first approached through what we call the index of thermophily ($I_t$). This index is based on the close correlation between the distribution of many plant species and temperature. Instead of relying on a single species, $I_t$ is calculated on the basis of all Arctic thermophilous species found in each km² of a given study area (Elvebakk 1990, Karlsen 1993, Joly et al. 2010). $I_t$ reflects summer temperature and can be considered more stable and robust than if a single species were used as an indicator of thermal conditions, since the occurrence of one species is more susceptible to random events. Most previous studies merely establish the strong positive correlation of $I_t$ with the sum of temperatures during the growing season at a given study site, e.g. Svalbard (Elvebakk 1990, Brosø & Elvebakk 2000), East Greenland (Karlsen & Elvebakk 2003), northern Norway (Karlsen et al. 2005).

Second, it is necessary to acquire knowledge of temperature distributions in order to identify any connection between temperature and vegetation. The climatic data available for the Arctic in general (Pape et al. 2009) and Svalbard in particular (Førland et al. 1997) are limited: there are only 5 official meteorological stations with climatic records going back in time. Because successful modelling of the spatial distribution of temperature requires a dense network of temperature-recording stations, detailed studies of relationships between plants and temperature on Svalbard cannot be made. To overcome this shortage of data, we established (for summer 2000) a field network of temperature-recording sensors on the northern side of the Kongsfjorden (for easy access). Joly et al. (2010) studied correlations between $I_t$ and different synthetic temperature distribution characteristics in the Kongsfjorden area (Svalbard). Because these preliminary results were based solely on temperatures recorded during a single summer (2000), they were not statistically robust. To strengthen confidence in the results and conclusions, the present study is based on 2 improvements: (1) more temperature data were collected through a field campaign conducted over a further 5 yr period (2001 to 2005); (2) in addition, more sensors were distributed more widely in the study area to ensure a better quality of temperature estimations. In the present study, we analyse the spatial relationship between $I_t$ and another synthetic temperature characteristic, commonly called ‘growing degree days’ (GDD) (Thorhallsdottir 1998, Parviainen et al. 2008) but also known as the ‘effective temperature sum’ when relating the temperature of boreal or Arctic environments to vegetation (Tuhkanen 1993, Karlsen et al. 2005).

This paper has 3 main objectives. (1) To establish an interpolation model based on measurements made in the field using 50 temperature sensors. This model should be specific enough for it to restore continuous surface raster summer heat at local scale. (2) To evaluate the capability of temperature values obtained from the model to assess the distribution of $I_t$ values. Solving a multiple equation gives us the opportunity to estimate temperature values. (3) To extrapolate surface temperatures from $I_t$ values to test the robustness of the model.

We describe how an accurate interpolation model was created that is capable of generating a continuous temperature raster surface representing temperature during the growth season (GDD) from spot
temperature measurements made in the Kongsfjorden area. Model performance is then evaluated using estimated temperature values to predict the distribution of \( I_t \). Taking the thermophily index as a thermo-indicator, we check the accuracy of the temperature model to determine how it may be used for spatial interpolation and extrapolation. Finally, the performance of GDD extrapolated (using \( I_t \)) beyond the northern part of the Kongsfjorden area where it was calculated is assessed.

2. DATA AND METHODS

The study was performed in the Kongsfjorden and Engelsk Bukta area, a small region located in northwest Spitsbergen, Svalbard (Fig. 1A). The study area is located in the transition zone between the open sea to the west and the ice-covered inland area to the east. \( I_t \) is subdivided into geographical sectors: Brøggerhalvøya to the south, Ossian Sars-fjellet to the east, and the northern coastal fringe from Dyrevika to Kapp Guisez (Fig. 1B). The surrounding landscape is made up of glaciated mountains, with some summits approaching 1000 m, whereas the seashore is mostly characterized by a rather broad flat area, the ‘strandflat’, along which raised glacio-marine terraces are segmented by glacier outwash cones. A permanent climatic station is located at Ny-Ålesund, where the mean July temperature is 4.9°C and annual precipitation is 385 mm (Førland et al. 1997).

This area was covered by observations of temperature and plants. The superposition of the 2 networks (see Figs. 1B & 3) is almost the same, except where loggers have been removed for technical reasons. All available data are presented in Section 2.1, including biogeographic data defining \( I_t \) categories, field temperature data and environmental data derived from a digital elevation model (DEM) and SPOT satellite data. Section 2.1.1 explains how interpolation techniques were applied to generate raster layers representing daily average temperature surfaces. Section 2.2.2 explains how the GDD is calculated and interpolated based on the daily average temperature raster layers. A flow chart in Fig. 2 summarizes all the technical steps in the study.

2.1. Data

2.1.1. Temperature data

Temperature data were recorded from June 2001 until August 2005 using thermal sensors (HOBO H8 Pro Temp/external logger, 64 Kb memory, with external thermocouple). The sensors are adequately protected against direct solar radiation and rain by small shelters, similar to but smaller than those used for measuring at standard meteorological stations. The openings on the circumference of the shelters provide good ventilation inside the box. Preliminary tests were carried out both in France and in Svalbard.
to assess the accuracy of the system with satisfactory results. The values obtained from the loggers were first compared with the ones from standard meteorological sensors, and the standard deviation of the differences never exceeded 0.3°C (Météo-France 2000). Comparisons were also made between all the loggers placed in their own shelter and placed in a thermally homogeneous room. The differences were recorded within a range of 0.5°C.

The daily average temperature for 460 d (92 d per summer × 5 summers) was calculated. Forty-five temperature loggers were distributed throughout the study area (Fig. 1B), positioned as evenly as possible to sample the whole range of elevations, all slope orientations and all landscape contexts (proximity to fjords or the sea, proximity to glaciers, stable tundra or moraines). The sensors were positioned 20 cm above ground level as close as possible to the plants because the main goal was to study temperatures in relation to the distribution of vegetation. The influence of ground albedo is low—even if measurements are made very close to the ground surface—because soil and the vegetation are mostly dark: solar radiation is essentially transformed into sensible heat. Some sensors did not operate for the whole period for various reasons, e.g. damage by bears, poor waterproofing, or battery failure at low temperature, but not all loggers could be replaced. Therefore there is no perfect overlap between the area sampled by the temperature sensors (Fig. 1B) and the area covered by the botanical observations (see Fig. 3).

Because the number of operational loggers varied from one month to another, for each of the 460 d, some 24 to 40 temperature values were available depending on the number of operative temperature sensors. By mid-August, after the devices had been serviced for another year, at least 45 temperature measurements per day were recorded. But gradually, as and when devices failed, the number of available daily records decreased. The minimum available daily measurements (24 to 28 depending on the year) were obtained in early July, just before the annual service of all loggers.

2.1.2. Floristic data

The study area was divided into 166 quadrats, each 1 km² (Fig. 3) defined by the WGS1984 datum and UTM map projection. Of these 166 quadrats, 122 were used for calibration because they are located in an area (the Kongsfjorden area) also sampled by temperature sensors, and 44 were used for validation (southern side of Brøggerhalvøya, Engelskbuuka). Each quadrat was thoroughly searched for a prede-
Nilsen et al.: Temperature and vegetation in the Arctic

fined group of vascular plants. Within each quadrat transects were placed in such a way that they covered all major habitat types, and plant recording took place along these transects. The species to be recorded were carefully selected based on previous knowledge of their temperature preferences and given an initial value reflecting their degree of thermophily.

2.1.3. Spatial database

Step 1. (Fig. 2) A spatial database was created to enable temperature interpolation (Joly et al. 2011), with 12 spatial grids (rasters) describing environmental conditions in the Kongsfjorden area. These variables were derived from a DEM model (Norwegian Polar Institute) and from a SPOT image (from 1991, Spot Image). All these data were resampled at a resolution of 50 × 50 m so they could be georeferenced on a common spatial grid.

Step 2. (Fig. 2) The local topography was described by 7 variables: (1) elevation (Elev; m); (2) slope gradient (Grad) of 0 to 90°; aspect (Aspe) broken down into (3) sine (sin-Aspe) and (4) cosine (cos-Aspe); (5) enclosure or exposure index (Expos): a deeply embanked valley or a narrow valley bottom takes a negative value whereas a high point (summit of a hill, crest line) has a positive value (Joly et al. 2012); and (6) rugosity (Rugo). A square 3 × 3 pixel-window was successively positioned on each pixel of the DEM. For each window, a local first-degree polynomial was calculated and provided a plane adjustment surface from which slope gradient and aspect were derived. For rugosity, the difference between the altitude value given by the DEM and the altitude value given by the polynomial adjustment was first calculated for each pixel of the local window. The standard deviation of these differences was then processed to give rugosity. 

(7) Solar radiation (SRad) is a major factor in plant growth. This parameter is a theoretical value calculated for a cloudless day. To estimate the best conditions of potential incoming energy, global solar radiation was given for the summer solstice (21 June). One value was calculated for each hour of the day, using both solar (height above the skyline and azimuth) and topographical characteristics (gradient and aspect) according to the equations given by Dumoulin & Parizet (1987). The shadow effect due to masking landforms was also integrated. Finally, the 24 values were summed to obtain a theoretical estimation of daily global solar radiation. It should be stated that the cases of thermal inversion are rare during summer in the High-Arctic because there is sunshine also at night (except for late August), and radiative cooling is very limited, even at the end of the nocturnal period.

Basic geoprocessing gave 4 more variables: distance (8) to the nearest ridge (DRidge), (9) to the nearest river line (DRivL), (10) to the open sea (DOSe), and (11) to the fjord (DFjo). Finally, SPOT channels 3 and 4 were used to calculate (12) the normalized deviation vegetation index (NDVI). Variables 8 and 9 often significantly explain the spatial variation of temperatures because valleys are sheltered from cold winds coming from the glaciers so that temperature is usually higher than on ridges exposed to wind.

Twelve variables in the analysis could be considered a large amount. However, before the analysis it was unknown which one would best explain the spatial variation of temperature. In fact, this was not a problem, because the method of analysis was designed to identify the leading explanatory variables, the others being omitted from the estimation calculation.

An array of cross-correlations between all pairs of variables showed that there was high collinearity between altitude and Rugo (r = 0.87) and DRidge (r = 0.92). Other variables (Grad, SRad, DFjo, NDVI) display much lower collinearity with elevation than previously (the proportion of common variance between these 4 variables and elevation is 30, 26, 27 and 19%).
respectively). The main problem is to determine the threshold beyond which the effects induced by collinearity may be detrimental (Belsley et al. 1980), even if Schroeder et al. (1986) asserted that there are no statistical tests for determining whether or not multi-collinearity is a problem. The occurrence of collinearity is evaluated here by the correlation coefficient, which is more intuitive than the 'condition index' normally used (Bressoux 2008). We set it at -0.55 (30% of common variance between pairs of collinear variables). Rugo and Dridge, which are the 2 main pairings with comparatively high collinearity, have been removed. So, only 10 variables remain in the subsequent interpolations without, for the reasons just mentioned, a partial correlation being performed.

2.2. Methods

2.2.1. Interpolation process

Step 3. (Fig. 2) Interpolation used both regression and kriging (Pape et al. 2009, Joly et al. 2011). A linear regression was applied to correlate the daily temperature average (dependent variable) from the recording stations and their 10 environmental characteristics (explanatory variables). All the significant explanatory variables at the 5% significant threshold (p < 0.05) were included in a multiple regression function in the form:

\[ y = a_1v_1 + a_2v_2 + \ldots + a_nv_n + \varepsilon \]

where \( y \) is the dependent (predicted) variable (temperature), \( a_1, a_2, \ldots, a_n \) are the regression coefficients, \( v_1, v_2, \ldots, v_n \) are the environmental variables (independent variables) significantly correlated with \( y \), and \( \varepsilon \) is the residual vector. We used the stepwise method (Hocking 1976) with the ‘forward selection’ for selecting the variables identified as significant by the previous test. Despite some drawbacks (Blanchet et al. 2008), this method is appropriate for solving collinearity between explanatory variables. By solving this equation, the ‘regressed’ temperatures are calculated and then the residuals are calculated for each station by cross validation (Stone 1974, Plutowski et al. 2006). Residual values correspond to the deviation between observed and estimated temperatures.

The residual values associated with the recording stations were kriged in order to account for local variations in temperature. The kriged residuals were then added to the regressed temperature to obtain the final estimated temperature. The parameters of both regression and kriging were used as cartographic operators for calculating a temperature value at each pixel of the reference grid.

As \( I_t \) values were not scaled initially with the same resolution (50 m vs. 1 km), an aggregation (resampling) process was applied: the average of the 400 primary pixels included in each of the 122 quadrats (1 x 1 km) was calculated.

2.2.2. Calculation of ‘growing degree days’ (GDD)

Step 4. (Fig. 2) GDD was used to describe the global summer temperature distribution. The interpolation procedure was applied to each daily temperature average during the period under consideration (92 d x 5 summers). Thus the data base was enlarged by 460 new raster data layers, each layer representing the spatial distribution of average temperature of one particular day during the recorded period. These raster layers were then processed by map algebra to yield GDD values. For each pixel in the raster layers, GDD was computed by summing the 460 average daily temperatures >0 (negative values were omitted from the calculation) and then dividing that total by 5 to obtain a summer average GDD. The data were then tabulated for statistical analysis. GDD values at 1 km² resolution were included in the interpolation process as explanatory variables for \( I_t \) (Step 8 in Fig. 2).

2.2.3. Calculation of \( I_t \)

Step 5. (Fig. 2) \( I_t \) was calculated for each 1 km² quadrat. Calculations were based on the thermophily value, frequency and dominance of each thermophilous species (Elvebakk 1990, Karlsen & Elvebakk 2003). All thermophilous plants were recorded on a frequency \( (f) \) scale from 1.0 (‘single locality’), 1.5 (‘rare’) to 2.0 (‘scattered or common’). ‘Single locality’ is defined as the occurrence of one or several individuals within a single area of 10 x 10 m or less. ‘Scattered or common’ is defined as occurrences within at least 5 such 10 x 10 m areas within a distance of at least 200 m. ‘Rare’ is intermediate between these 2 definitions. A dominance \( (d) \) scale involves those already recorded as ‘scattered or common’ if, in addition, they occur in certain quantities. A value of 1.5 is given to ‘subdominants’ with 20 to 50% cover in the 10 x 10 m areas, and 2.0 for ‘dominants’ with cover values >50%.

In this study thermophilous species are defined and categorized into 3 groups: T4, T5 and T6. T-numbers indicate the average July isotherms on a circumpolar
scale that best fit with the given plant species’ northernmost distribution. For each group an associated relative thermophily value \( T_i \) ranging from 1.0 (T4), 1.5 (T5) to 2.0 (T6) is assigned to the given species. A more thorough list of thermophilic Arctic species and their grouping can be found in Elvebakk (1990) and Karlsen & Elvebakk (2003). The values of all \( n \) species of thermophiles in each of the 122 quadrats were included in the following formula for \( I_t \):

\[
I_t = \sum T_i (f + d)
\]

(2)

\( I_t \) values are thought to reflect the temperature climate near the ground (below and above). \( I_t \) also reflects the temperature sum during the growing season, and as plants are sessile organisms, it also represents a climatic response to temperature conditions covering a long period of time. \( I_t \) values at 1 km\(^2\) resolution were introduced in the interpolation process (Step 6 in Fig. 2) as explanatory variables accounting for temperature (GDD).

2.2.4. Validation of estimated values

There are 2 ways to assess the ability of GDD to estimate \( I_t \) and vice versa. The first one is to examine the residuals obtained by using a cross validation process. This is only possible for the Kongsfjorden area where both types of data are available. The second one involves extrapolating GDD to the Engelskbugta area (using Eq. 1, as \( I_t \) is known there; section 3.1) and then estimating \( I_t \) using the GDD parameter (section 3.2). The results of this estimation of \( I_t \) can then be compared with the 44 observed \( I_t \) values.

3. RESULTS

There are 3 main results. The model resulted in the interpolation of GDD and \( I_t \) and the calculation of residuals to assess the quality of estimates.

3.1. Spatial structure and map of GDD

Step 6. (Fig. 2) The method used to interpolate daily temperatures (Section 2.1) is also applied to GDD, which thus becomes the predicted variable. The explanatory variables are provided by the 10 spatial gridded data related to environment previously described to which we add \( I_t \). Of these 11 variables, only 6 are recognized as significant at the 5% level by tests (Table 1a).

<table>
<thead>
<tr>
<th>( \text{Elev} )</th>
<th>( I_t )</th>
<th>( \text{DRivL} )</th>
<th>( \text{NDVI} )</th>
<th>( \text{Expos} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDD</td>
<td>–0.93</td>
<td>0.45</td>
<td>0.43</td>
<td>–0.42</td>
</tr>
<tr>
<td>( \text{DOSe} )</td>
<td>GDD</td>
<td>( \text{DRivL} )</td>
<td>( \text{DFjo} )</td>
<td>cos-Aspe</td>
</tr>
<tr>
<td>( I_t )</td>
<td>0.71</td>
<td>0.38</td>
<td>–0.36</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Elevation is a very powerful factor explaining the spatial variation of GDD (Table 1a). The remaining 4 significant variables are related to vegetation (\( I_t \) and NDVI) and topography (\( \text{DRivL} \) and \( \text{Expos} \)). The GDD model stemming from regression gives:

\[
\text{Estimated GDD} = 484.1 + (–0.30 \text{Elev}) + (12.8 I_t) + (0.12 \text{DRivL}) + (0.31 \text{NDVI}) + (–0.36 \text{Expos})
\]

(3)

Notice that NDVI has \( r = –0.42 \) and the coefficient of regression of NVDI in Eq. (3) is 0.31 (positive). This is possible due to the partial collinearity that links elevation and NDVI (section 2.1.3.). It is known that among the consequences of statistical collinearity between the explanatory variables, the sign of their parameters may be changed (Foucart 2006).

This index is called ‘estimated GDD’ to distinguish it from GDD (that is the total sum of observed temperatures for all 460 d; section 2.2). By applying Eq. (3), estimated values can be calculated for every point of the study area and a continuous field of GDD mapped. Because \( I_t \) (the 11th explanatory variable) is available for the Kongsfjorden and the Engelskbugta areas, the GDD map covers both areas with the same accuracy. The map (Fig. 4) shows the strong influence of the negative elevation gradient. Temperature decreases by 30° GDD per 100 m. The second main point to be highlighted is the influence of continentality. The highest GDD values are concentrated close to the heads of the fjords.

Step 7. (Fig. 2) Comparing the estimated values with the observed values provides a residual value that yields information about the quality of the estimate. Residual values from the interpolation process are assessed by cross validation. The standard deviation of estimated GDD is 26 (GDD observed values are within the 220–538 range). The scatterplot of the interpolated and estimated GDD is shown in Fig. 5.
The quality of the model is high (the coefficient of determination \( r^2 \) for estimated GDD is 0.85).

### 3.2. Spatial structure and map of \( I_t \)

**Step 8.** (Fig. 2) \( I_t \) is estimated from the same 10 explanatory variables as before (sections 2.1 and 3.1) to which we add GDD (in °C). The spatial structure of \( I_t \) was tested using the same protocol as for calibrating the spatial structure of GDD.

The coefficient of correlation associated with each variable shows that \( DOSe \) \((r = 0.71) \) and climate expressed as GDD \((r = 0.38) \) are 2 powerful factors in explaining the spatial variation of \( I_t \) (Table 1b). \( DRivL \) and \( \cos-Aspe \) describing the appearance of the slopes account for a significant amount of the variation of \( I_t \). The greater the distance to the nearest river line, the lower the \( I_t \). The proximity of a river line sheltered from winds is favorable to plant growth. The south-facing slopes (cosine = -1) are associated with rather high values of \( I_t \), while the values of \( I_t \) decrease if the slopes are oriented to the north (cosine = 1). Winds from the south are warmer than those from the north and promote plant growth. What is highlighted through the north–south orientation of the slope is in fact more the sector from which the wind blows than the radiation received at the surface. \( SRad \) has little explanatory value \((r = 0.16) \). \( Elev \) has only a slight influence on \( I_t \) spatial variation \((r = -0.05) \) because \( I_t \) quadrats are all located on the strandflat area. The 6 significant variables are used for establishing the following regression:

\[
I_t = 57 + (0.075 \times DOSe) + (0.11 \times GDD) + (0.003 \times DRiveL) + (0.06 \times Dfjo) + (-0.03 \times \cos-Aspe) + (0.13 \times Grad)
\]

The calculation of \( I_t \) is based on the frequency of plants that have been observed in the Kongsfjorden and Engelskbukta areas by botanists. Because \( I_t \) is a function of 6 variables (Eq. 4), including GDD, which was observed only in the Kongsfjorden area, the spatial structure of \( I_t \) including the GDD explanatory variable has been analysed for the Kongsfjorden area only. Using the spatial distribution of GDD calculated by interpolation only (estimated GDD; Eq. (3)), it is possible to estimate \( I_t \) for the 2 areas. However, the \( I_t \) estimation is strictly valid north of the dashed line only (Fig. 6). In Engelskbukta (south of the dashed line), where GDD has been extrapolated, \( I_t \) is more hypothetical.

Within the Kongsfjorden study area, high \( I_t \) values are located mainly on the south-facing slopes but a gradient from west to east is also revealed with lower values to the west (Kvadehuk and Kapp Guissez) and higher values to the east (Ossian Sars-fjellet). \( R^2 \) for estimated \( I_t \) is 0.68. The highest \( I_t \) values affect the head of the Kongsfjorden and especially Ossian Sars-fjellet. Engelskbukta is located on the southern slope of Brøggerhalvøya and so the \( I_t \) values are higher than for the northern slope.

**Step 9.** (Fig. 2) The standard deviation of residuals is 9 \((I_t \) observed values range from 0 to 81). More than 57 % of residuals are within the limits -5 to +5. Fig. 7 shows the correlation between observed and estimated \( I_t \). Quadrat values are mostly close to the re-
gression line, revealing a relatively good fit of the model ($R^2 = 0.52$). The few high residual values are concentrated in Ossian Sars (–41, –17), which is known to be a hot spot in Kongsfjorden.

### 3.3. $I_t$ estimated values in Engelskbukta

**Step 10.** (Fig. 2) The GDD estimation is established with data from temperature loggers located in Kongsfjorden only. However, Eq. (3) enabled us to restore GDD from observed $I_t$ values in the Engelskbukta area. Therefore, when applying Eq. (3) using extrapolated GDD to a wider area including Engelskbukta, it is interesting to compare the observed and estimated $I_t$ values to see whether the model works equally well in this area, which is not taken into account for calibrating $I_t$. The results of this estimation can then be compared with the 44 observed $I_t$ values. It confirms that the model is also suitable for Engelskbukta (Fig. 8). The $r^2$ of the model is >0.65. The standard deviation of residuals is 4.8 and the deviating residuals vary mostly between –11 and +11. The error is even smaller than in Kongsfjorden because Engelskbukta is more homogeneous.

### 4. DISCUSSION

#### 4.1. Grid cells

The aggregation of botanical observations within 1 km$^2$ grid cells makes it possible to fully encompass the minimum area of plant associations. Some cells have unequal habitat diversity. Karlsen & Elvebakk (2003) introduced a modified version of $I_t$ where unequal habitat diversity was compensated for. This was not done here. However, modifications have moderate influences when the grid cells are as large as 1 km$^2$, as these normally will include both ridges, intermediate areas, snowbeds and concave landforms supporting mires. In short, it means that since the plants are observed within 1 × 1 km squares,
these squares will normally include all major habitats in the region. Thus, the method is independent of the mosaic-like distribution of true surface temperatures, which are very different in wet mires and depressions with long snow duration as compared to dry ridges.

The temperature values, recorded by sensors scattered in space, are interpolated on grids of 50 m resolution. The problem of heterogeneity of cells, especially in mountain areas where the topography can be very rugged, can also arise here (Tveito & Førland 1999). This problem occurs regularly in the interpolation process and there is no real solution. This is actually a problem of scale. It must be admitted that a cell is a ‘black box’ whose internal heterogeneity is inaccessible, and this can lead to estimation errors if the sensors distributed at random. In the present study, this is not the case, however, since they were installed in homogeneous areas, representing a wide area around them.

### 4.2. Factors influencing $I_t$

This work is mainly focused on assessing the co-variation of temperature and plants. Temperature taken into account here was recorded by means of sensors located 20 cm above the ground to measure the temperature of the air layer located immediately above the soil and vegetation cover. Plant temperature is certainly different. In addition, plant growth depends not only on the surface temperature. It also depends on the temperature of the surface layer of the soil, which could exert a strong control on plant growth. Lupi et al. (2012) have conducted some experiments in boreal ecosystems on the relationship between soil temperatures and plant growth (see also Van Cleve et al. 1990, Strömgren & Linder 2002, Körner 2006). It was shown that cambium phenology in trees is better explained by air than soil temperature. We can suppose that what is true for boreal trees is true also for the Arctic tundra. Another work (Monteiro et al. 2011, p. 395) reveals that the foliage of tussock grass in the Andean Altiplano operates ‘near air temperature, while the center is occluded with debris that insulates the root-stock and the meristems therein from both, radiated warming and cooling’. Microclimatological measurements carried out in Svalbard also showed that the soil temperature is strongly correlated with that of the air, even if the daily temperature amplitude falls because of the high insulating power of the plant matter (Joly 1994). All these observations tend to show that the temperature of the air layer located 20 cm above the ground reflects the thermal constraints faced by plants.

### 4.3. Factors explaining GDD and $I_t$ spatial structure

The first point to be highlighted is the influence of 2 main factors on GDD estimated values. DOSe as a continentality factor is not introduced into Eq. (3). However, Fig. 4 shows that most GDD highest values (>399°C) are concentrated close to the heads of the fjords, far from the open sea. It is likely that this effect is statistically captured by $I_t$ which is itself strongly dependent on DOSe ($r = 0.71$; Table 1b). Similarly but reversed, altitude is not included in Eq. (4). We know that elevation is a powerful factor limiting the growth and survival of plants (Bell & Bliss 1980, Wielgolaski & Karlsen 2007). Here, GDD, which is closely associated with elevation ($r = -0.93$, Table 1a), probably captures the effect of elevation on the distribution of plants and thus $I_t$. However, because most quadrats are located at low altitude along the strand-flat, the potential influence of altitude on $I_t$ is reduced. Still, thermophilous plants decline markedly with altitude, although the very steep mountain slopes in the present study area makes it less suitable for analyses of the altitudinal gradient than, say, the study from east Greenland by Karlsen & Elvebakk (2003) using the same method. The influence of vegetation on the spatial distribution of temperature is also visible through the dark patches corresponding to the areas covered by stable tundra (Fig. 4). This is consistent with Joly & Brossard (2007).

### 4.4. Deviation between observed and estimated values

$I_t$ is calculated for each quadrat (1 km²); however, for some of them, thermophilous plants might be concentrated in a small sector only. In such cases, the estimation fails because $I_t$ does not reflect the environmental characteristics of the entire quadrat in question. This is especially the case for Ossian Sarsfjellet where the environment is very heterogeneous at the 1 km ground resolution. There, the quadrats include a narrow belt of fertilized soil along the base of a bird cliff, close to the fjord, where plants grow and the upper part of the cliff (>200 m elevation). The study by Joly et al. (2010), based on recordings made during a single summer (2000), noted the difficulties in interpolating $I_t$ values at this site. An improved 5 yr
record of temperature data did not enhance the result in this respect. This particular situation may explain the marked deviation between observed and estimated values. Smaller \( I_t \) recording units, i.e. 500 × 500 m quadrats, would be more suitable when the landscape structure is heterogeneous. Such resolution of observation units should be in better agreement with temperature data.

Where glaciers have retreated significantly, the landscape is dominated by young moraines colonised by pioneer plants that give a low observed \( I_t \) value with negative residuals. Elsewhere, the results are better when habitat characteristics offer a good potential for thermophilous plants considering Eq. (4): short distance from the sea, high rugosity, mid-range values for temperature sum and NDVI. In 2 cases where the observed \( I_t \) values reach 75 and 81, the residuals are positive (34) and very high. The sites in question are also located close to the sea, but with very local landscape conditions resulting in sheltered habitats with very favourable temperature conditions, strengthening the effect of the west–east thermal gradient. Both these quadrats include heterogeneous areas where altitude differs by >100 m, which contributes to minimizing the value estimated by the model. Under such circumstances the 1 km ground resolution of the \( I_t \) data recordings is unsuitable and largely explains the model’s poor performance.

Correlation between \( I_t \) and temperature may be locally affected by geology (Elvebak 1982, Brossard et al. 1984). The Kvadehuken strandflat is covered by well-drained limestone gravel and stone deposits that accommodate a limited number of vascular plants. In such circumstances the \( I_t \) values are low, close or equal to 0. A modified version of the index would have compensated for this low habitat heterogeneity. However, a 30% increase in already low values would not yield a significantly different result. The conclusion is that this cold and windswept coastal area does have a very low temperature regime that is reflected in \( I_t \) values even if the geology of the area has particular features (it should be noted that this ‘inversion’ is not radiative as in the case of conventional temperature inversions). Mountainous areas are also characterized by frequent high temperatures because of the regular occurrence of temperature inversions. During such weather conditions, the warmest places are located in the interior part of the mountains. The altitudinal gradient is ~1° per 100 m (Brossard et al. 2002). The highest values are located on the well-exposed south-facing slopes at 500 m elevation. These thermal conditions might seem favourable to thermophilous plants. However, these habitats are also characterized by a short growing season and frost becomes increasingly frequent with elevation (Crawford 2008, Körner & Alsos 2009). On the whole, plant growth and survival is drastically restricted here.

5. CONCLUSIONS

The thermal data layers created at fine-scale resolution (50 × 50 m) from HOBO temperature logger measurements reflect temperature distributions in areas with wide environmental and topographic variability. The index method defines a broad habitat diversity within each study unit, and is independent of mesoscale surface temperature contrasts, which can be very high between a wet mire and an adjacent dry ridge. Still, the resolution of the sampling units of \( I_t \) is large (1 × 1 km), and integrated \( I_t \) values from such sampling units yield a weaker correlation with measured temperatures when the units contain highly heterogeneous environments, such as near the Ossian Sars-fjellet bird cliffs, or highly homogeneous areas, e.g. on monotonous steep mountain sides. Statistical relationships between botany and climatology were identified and quantified through the intersection between \( I_t \) and GDD, both of which were observed in the Kongsfjorden area. These statistical relationships were then applied to another area (southern side of Brøggerhalvøya) where GDD can be estimated from botanical data only.

Despite some divergence, comparison and correlations between \( I_t \) and thermal data sets yield very encouraging results. Three contrasting areas can be recognized: (1) western areas close to the open sea characterized by broad expanses covered by limestone pebbles and a pure maritime polar climate, neither of which favour thermophilous plants; (2) eastern, inland parts of Kongsfjorden on average 3°C warmer with a high frequency of GDD up to 400°C and very favourable to thermophilous species; (3) mountain areas characterized by frequent high temperatures, but frequent frosts on the whole drastically restricting plant growth.

The model as it is established for the Kongsfjorden area might be applied by extrapolation to other places on Svalbard. The low error of estimation when applying it to Engelskubukta tends to confirm this assumption. Theoretically and technically, the model can be applied anywhere where the environmental variables and climatic indices used in Eq. (3) are known. As DEM is available for the whole of Svalbard, DOSe, Aspe, and Rugo can be calculated. The
source for processing NDVI is obtained from satellite (SPOT) images. However, an essential environmental variable such as GDD data was available from the Kongsfjorden area only, and application of the model is therefore strictly restricted to this region and its immediate vicinity. Further studies are needed to test whether the model can be exported to other areas of Svalbard where the regional climate and fjord lengths are different.

The first phase of our project (evaluating changes in vegetation on Svalbard due to climate change) has been completed. The close connections between the 2 indices crossed (GDD and $l_t$) make it possible to consider potential future developments. Thus, insofar as the distribution of plants is well known throughout Svalbard, it will be possible to infer the spatial variation of GDD. That will be the next step in the project, but precisely for the reason outlined above.

LITERATURE CITED


Danby RN, Hik DS (2007) Evidence of recent treeline dynamics in Southwest Yukon from aerial photographs. Arctic 60:411–420


Joly D, Nilsen L, Brossard T, Elvebakk A (2010) Plants as bioindicator for temperature interpolation purposes: analyzing spatial correlation between botany based index of...
thermophily and integrated temperature characteristics. Ecol Indic 10:990–998


Submitted: November 23, 2012, Accepted: June 16, 2013

Prooeds received from author(s): October 23, 2013

Editorial responsibility: Tim Sparks, Cambridge, UK